Assignment - 2

Name	:	Md. Miju Ahmed
ID	:	2010676104
Session	:	2019-2020
Course	:	Neural Network and Deep Learning
Course Code	:	CSE4261
Dept	:	Computer Science and Engineering
Date	:	27-05-2025

Below shows the activation function list for every model that they were used for building their architecture and also shows the accuracy after using the different activation functions in the head:

Model	Activation Function Used	Accuracy after using ReLu in head	Accuracy after using Softmax in head
MobileNetV2	ReLu6	15.95%	69.80%
ResNet50	ReLu	6.45%	5.65%
VGG16	ReLu	8.95%	25.80%
EfficientNetB0	Swish	5.00%	5.00%
DenseNet121	ReLu	9.20%	66.05%
NASNetMobile	ReLu	24.05%	72.85%
EfficientNetV2B0	Swish	5.00%	5.00%
InceptionV3	ReLu	28.60%	74.85%
Xception	ReLu	47.25%	76.70%
InceptionResNet V2	ReLu	28.50%	79.30%

Here we can see that, if we use the activation function as a ReLu without softmax then the accuracy are decreased for those models.

Below shows the list of the CNN models those are used regular kernel, deformable kernel, dialated kernel, depthwise separable kernel, modified depthwise-separable kernel, and pointwise kernel

Kernel Types	**CNN models**
Regular kernel	VGG16, ResNet50, DenseNet121
Deformable kernel	Deformable ConvNet v1, Deformable ConvNet v2, YOLOv4
Dialated kernel	DeepLabv3, WaveNet, ESPNet
Depthwise separable kernel	MobileNetV1, Xception, EfficientNet
Modified Depthwise-Separable kernel	MobileNetV2, MobileNetV3, FBNet
Pointwise kernel	InceptionV3, ResNet (Bottleneck), MobileNet series

My chosen CNN model is InceptionResNetv2. Below breaking down the feature map evolution layer by layer for this architecture, focusing on how each layer transforms the input into higher-level representations.

1. Stem Block

- Input: 224×224×3 (ImageNet input size)
- Layers:
  - 3×3 Conv → 32 filters → 149×149×32
  - 3×3 Conv → 32 filters → 147×147×32
  - 3×3 Conv → 64 filters → 147×147×64
  - MaxPool + Conv → 73×73×192

Feature maps: Capture edges, gradients, color blobs. Low-level features.

2. Inception-ResNet-A Block (5x)

- Output: 35×35×320
- Each block contains:
  - Parallel branches (1×1, 3×3, 5×5 convolutions)
  - Concatenation + Residual shortcut

Feature maps: Learn mid-level patterns — corners, textures, basic shapes.

3. Reduction-A Block

- Downsamples to 17×17×1088

Feature maps: Begin learning object parts like heads, wheels, leaves.

4. Inception-ResNet-B Block (10x)

- Output: 17×17×1088
- Uses narrower filters (like 1×7, 7×1) to capture asymmetrical patterns.

Feature maps: Represent larger structures — faces, windows, animal parts.

5. Reduction-B Block

- Output: 8×8×2080

Feature maps: Now abstract enough to represent full objects — cars, dogs, etc.

6. Inception-ResNet-C Block (5x)

- Output: 8×8×1536

Feature maps: Final object-level concepts before classification.

7. Final Layers

- Global Average Pooling → 1×1×1536

- Dropout + Dense Layer → Softmax output (e.g., 100 for CIFAR-100 subset)

Feature maps: Fully condensed object understanding — one feature per class.